RESPONSE TO REVIEWER #2 FOR *GEOSCIENTIFIC MODEL DEVELOPMENT*:
MANUSCRIPT EGUSPHERE-2024-1456
BY SEUNG H. BAEK, PAUL. A. ULLRICH, BO DONG, AND JIWOO LEE

We thank Reviewer #2 for thoughtful and constructive feedback. This Response to the Reviewer file provides a complete documentation of the changes that have been made in response to each individual comment. Reviewer's comments are shown in plain text. Authors' responses are shown in **bold**. Quotations from the revised manuscript are shown in ***bold italics***.

Reviewer #2

Thank you for inviting me to review the paper: "Evaluating downscaled products with expected hydroclimatic co-variances" by Baek et al.

First, please accept my apologies for the slowness in returning this review.

This paper is important and interesting because rather than just considering climate variables of concern in isolation (e.g., their role in extremes), the manuscript emphasizes generating the correct covariances between quantities when downscaling. On that basis, I hope the manuscript will eventually appear in print in some format.

The two obvious variables needing consistency are temperature and heavy precipitation, as noted in the Abstract.

Below are some suggestions that may help towards creating a revised paper version.

In the Abstract, please state the data used to test the two downscaling methods.

We now state that we assess LOCA2 and STAR-ESDM *"…as compared to European Centre for medium-Range Weather Forecasts Reanalysis v5 (ERA5) and two observation-based data products (Livneh and nClimGrid-Daily)."*

In the Abstract, please also provide a typical spatial scale of the analysis that data and models (e.g., RCMs) can currently achieve.

We now indicate that presently available downscaling products have ~5-km grid spacing. Further work is needed to determine the effective of resolution of the products given that they are derived from meteorological stations which are often more than 5 km apart.

Additionally, the Abstract needs to be easy for a wide audience to understand. The sentence "…our results suggest that statistical downscaling techniques may be limited in their ability to resolve non-stationary hydrologic processes as compared to dynamical downscaling" is slightly ambiguous. Please make clear in the abstract that the word "dynamical" implies a more process-based approach. And "non-stationary"—is that referring to climate change? An Abstract should be broadly understandable in isolation from reading the full manuscript.

We have revised the sentence to *"…our results indicate that statistical downscaling techniques may be limited in their ability to represent future physical processes that deviate from their historical behavior due to climate change."* We also indicate in the abstract that dynamical implies a more process-based approach and have removed the word "non-stationary" to make our abstract more broadly understandable.

The Introduction is good, as it clearly differentiates between statistical and dynamical downscaling. Around line 54, it might help to give citations to example meteorological datasets that are used at the very fine scale, and of course as validation.

We now cite Livneh et al., 2015 and Durre et al., 2022 as example datasets where suggested.

The authors could expand a little more around line 61 regarding how there is no guarantee that statistical relationships between variables identified for current levels of atmospheric greenhouse gases will remain valid for future higher GHG concentrations. However a good start is testing performance for current GHG levels.

We now expand on what was formerly Line 67: *"Additionally, statistical downscaling assumes that observed functional relationships will be preserved in the future (i.e., the stationarity assumption) despite climate change (Milly et al. 2008); however, there is no guarantee that historically derived statistical relationships will remain valid in the future."*

The paragraph starting "We compare" (line 117) is very helpful and differentiates well between dynamical downscaling (i.e. nested "RCMs", which contain physical process knowledge that is hopefully also valid for higher GHGs) and inferences from high-resolution meteorological data. The authors make good use of Tables (Table 1, CMIP6 models assessed and Table 2, "data" versus RCM). However, would it be possible to list in a Table, also measurement datasets. A combined Table might help.

We now list the observation-based datasets (ERA5, Livneh dataset, and nClimGrid dataset) in Table 1 and provide their respective grid spacings.

A further advantage of a carefully consider Table would be an opportunity to describe all the project names, ESM names, RCM names, datasets. I found myself continuously jumping between diagram captions and different parts of the text to fully understand what I was actually looking at. Some are not even defined – CONUS I guess is C for Canada, US for US? At first, reading the caption to Figure 1, I thought CONUS might be a dataset I had not heard of.

We thank the Reviewer for this suggestion. We have combined all previous tables into a single table to make it easier for readers to reference datasets. Table 1 now lists raw CMIP6 GCMs, the observation-based datasets, the statistically downscaled datasets, and the NA-CORDEX dynamically downscaled datasets. We now clarify in the first instance of the phrase "contiguous US" that we will herein refer to it as CONUS.

In isolation, Table 2 does not make full sense? If I've understood correctly (around line 125), then this Table is which RCMs (top row) are nested in which ESMs (left column). Correct? But I do not fully understand why ERA-Int is mentioned in the left column – surely the comparison would be between ERA-Int and each ESM/RCM combination? Maybe that is what the top row is getting at with "Analyzed" as each entry?

In addition to the ESMs (left column), NA-CORDEX downscales the ERA-Interim dataset across all of the RCMs. A key benefit of this is that it allows for comparisons of RCM biases across a common dataset. For instance, if different ESM/RCM combinations were to be compared against each other, it would be difficult to assess the extent to which the differences are due to ESMs or RCMs. We now clarify this in the caption: "*In addition to global climate models, NA-CORDEX downscales ERA-Interim (top of left column) across different regional climate models (this allows for a comparison of downscaling across a common dataset).*"

The aspect I like best about the analysis is the ability to compare the first column of Figure 1 (composite behaviour around a convective event) and Figure 2 (composite behaviour around frontal precipitation). To say the obvious, the curve shapes are very different. And critically, the relationship between the T and Precip curves is markedly different between the two. Please make sure that the reader is drawn to these differences in the main writing in the text (and/or discussion).

We now draw the reader towards details describing features of temperature and precipitation curves (as well as their spatial expressions) during frontal precipitation events with a key sentence near the beginning of the paragraph starting around what was formerly Line 177: *"Our selection of frontal precipitation events show a very different relationship between temperature and precipitation as compared to convective precipitation."*

The next and important comparison, which again needs to come out very clearly from the paper is how actual datasets perform when compared to nested RCMs? Which Figure should I be looking at to compare against Figure1 and 2? Figure 4 is raw CMIP6 GCM data – so not the nested RCM outputs? ESM/RCM combinations are our best estimate of fine-scale meteorological behaviours, and so it should be them that are being tested the most against data?

The main message of the paper is that statistical downscaling does not capture co-variances between P and T as well as ESM/RCM i.e. process (or dynamical) downscaling? So again, the reader needs to see that really clearly too – which I guess implies comparing Figure 1 and 2 against Figure 9?

The two comments above are related so we address them both here. As noted by the Reviewer, Figure 4 is the raw CMIP6 GCM results. Figure 10 is the ESM/RCM combination results that readers should compare to Figure 1 and 2 (as well as Figure 4). Our manuscript previously may not have been sufficiently clear to easily allow readers to compare across these figures. We have made concerted efforts to improve both the writing and quality of figures.

The reader needs to be steered more clearly, leading to the four main messages:

- How do convective P-T relationships differ from frontal P-T relationships?
- Do statistical downscaling methods fail at this, as they cannot differentiate between the two necessarily.
- Do GCM/RCM combinations capture these differences better? And if so, then:
- What do GCM/RCM combinations project into the future?

We thank the Reviewer for these comments. We have made concerted efforts to improve the quality of our writing and have made edits throughout the manuscript to more clearly steer readers towards our main conclusions. For instance, we now put in more references to specific figures in our Results section to more explicitly guide readers towards relevant results that support our key arguments.

The presentation is poor in places. For instance, focusing on Figure 5, the left-column x-axis label needs to be marked as “days”. Similarly, units are needed under the colourbars. The maps can be stretched by removing the white space of Canada and Mexico – this will make them easier to interpret. In many places, the diagrams can be enhanced in their appearance.

We now mark units as suggested and have revised our figures to remove more of the white space of Canada and Mexico. We have also worked to improve the quality of the figures more broadly. With these changes, we believe that the figures are easier to interpret for readers.

I am very happy to see the manuscript again, and in the meantime, I hope some of the suggestions and comments above are helpful.

We thank the Reviewer for constructive comments that have helped improve our manuscript.